◇ Monograph Excerpt ◇

Matter Antimatter Fluctuations

search, discovery and analysis of Bs flavor oscillations

Nuno Leonardo

Complete work published as:
Analysis of B_s oscillations at CDF, MIT Thesis (2006)
Matter antimatter fluctuations, Monograph, LAP Lambert (2011)

Chapter 11

Mixing significance

An analytical derivation of the expected amplitude significance of an oscillation signal is provided, based on relevant sample characteristics, and used for sensitivity estimations. The probability that the observed oscillation signal could be mimicked by a statistical fluctuation in the data is evaluated.

11.1 Formalism

In this section we explore an analytical characterization of the mixing analysis. A powerful mathematical technique for the analysis of periodic signals is the Fourier transform. The amplitude method employed for the search and the study of B_s oscillations is itself motivated and shares several advantages associated with the latter techniques, as well as with the thorough likelihood characterization of the data sample components. Effects of detector resolutions, partial reconstruction, and biases on the proper time distribution of the samples are explored; see also the expositions in [33, 96, 97]. The amplitude significance of a mixing measurement is derived.

11.1.1 Likelihood approach

The basic likelihood description in proper decay time space for flavor tagged candidates has the following form

$$f_\xi(t) \;=\; \frac{N}{2}\left(f_s\frac{1}{\tau}e^{-\frac{t}{\tau}}\theta(t)[1+\xi\mathcal{D}\cos(wt)]+(1-f_s)(1+\xi\mathcal{D}_B)B(t)\right) \qquad (11.1)$$

where N stands for the number of tagged events in the sample; f_s the fraction of signal events; $\mathcal{D}$ and $\xi \in \{-1,+1\}$ the flavor tagging dilution and decision; τ and w the lifetime and the oscillation frequency of the B system. The background model contains a possible

small flavor tagging asymmetry, which is assumed to be time independent; it is accordingly described by the asymmetry parameter $\mathcal{D}_B$, along with an empirical t distribution, $B(t)$, characterized by a decay constant τ_B.

The flavor asymmetry may be obtained as the difference of the expressions for the two tagging decisions

$$a(t) \;=\; f_+(t) - f_-(t) = N\left(f_s\mathcal{D}\frac{1}{\tau}e^{-\frac{t}{\tau}}\theta(t)\cos(wt) + (1-f_s)\mathcal{D}_B B(t)\right) . \qquad (11.2)$$

We shall focus in this section on the signal component. We assume the background may be effectively subtracted via additional likelihood factors describing discriminating quantities such as the candidates' mass along with others. Effects of detector resolution, trigger, selection and partial reconstruction are not included in the above expressions.

Detector resolution

The reconstructed proper decay time is given by

$$t \;=\; \frac{LM}{p}$$

where L and p are the (transverse) decay distance and momentum, and M the mass of the B meson. The uncertainty is generally given by

$$\sigma_t \;=\; \frac{L}{p}\sigma(M) \oplus \frac{M}{p}\sigma(L) \oplus \frac{\sigma(p)}{p}t$$

where σ denotes the partial uncertainties and the $\oplus$ symbol indicates that the terms combine in quadrature. The first, which is proportional to the uncertainty on the B hadron mass, is negligible; in fact, the average world value for M is taken rather than the reconstructed candidate's mass. The uncertainty in the decay vertex position is to first order independent of the decay length itself, while the momentum resolution deteriorates with increasing momentum. We assume for the moment that the absolute uncertainty on the decay length and the relative uncertainty on the B momentum are Gaussian.

The smearing effect on the proper decay time distributions (11.1) may be expressed as

$$f_\xi(t) \;\mapsto\; \int_{-\infty}^{+\infty} f_\xi(t')\mathcal{R}(t',t)\,dt' \qquad (11.3)$$

where $\mathcal{R}$ is the resolution function. The latter can be written as

$$\begin{aligned}\mathcal{R}(t',t) \;&=\; \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma(p)}e^{-\frac{(p-p')^2}{2\sigma_p^2}}\frac{1}{\sqrt{2\pi}\sigma(L)}e^{-\frac{(pt-p't')^2}{2M^2\sigma(L)^2}}(\frac{p'}{M})\;dp \\ &\simeq\; \frac{1}{\sqrt{2\pi(\sigma_l^2+\sigma_p^2t^2)}}e^{-\frac{(t't')^2}{2(\sigma_L^2+\sigma_p^2t^2)}}\end{aligned}$$

where we have assumed $\sigma(p)/p' \ll 1$ and defined

$$\sigma_l \equiv \frac{M}{p'}\sigma(L) \simeq \frac{M}{p}\sigma(L) \qquad \text{and} \qquad \sigma_p \equiv \frac{\sigma(p)}{p'} \simeq \frac{\sigma(p)}{p} \; .$$

We may write

$$\sigma_t = \sigma_l \oplus \sigma_p \, t \; . \tag{11.4}$$

In addition to the resolution smearing, we may include the biasing effects in the proper decay time distribution induced by trigger and signal selection criteria. These are described by the t-efficiency function $\mathcal{E}(t)$, presented in Section 5.3.1. This function is defined in terms of the *reconstructed* proper time, and is thus applied after the detector resolution smearing.

The integration in (11.3), with the biasing effects incorporated, identically to (7) and (8), for the signal terms, is given by

$$\begin{aligned}
&\int_0^\infty \frac{1}{\tau} e^{-\frac{t}{\tau}} \left(1 + \xi\mathcal{D}\cos(wt)\right) \mathcal{R}(t',t)\, dt' \cdot \mathcal{E}(t) \\
&= \frac{1}{2\tau} e^{-\frac{1}{\tau}(t-\frac{\sigma_t^2}{2\tau})} e^{-\frac{\sigma_t^2 w^2}{2}} \cdot \mathrm{Re}\left\{ e^{-iw(t-\frac{\sigma_t^2}{\tau})} \,\mathrm{Erfc}\left(\frac{\sigma_t^2 - t\tau}{\sqrt{2}\sigma_t\tau} + i\frac{\sigma_t w}{\sqrt{2}}\right)\right\} \cdot \mathcal{E}(t) \\
&\simeq \frac{1}{\tau} e^{-\frac{t}{\tau}} \left(1 + \xi e^{-\frac{\sigma_t^2 w^2}{2}} \mathcal{D}\cos(wt)\right) \cdot \mathcal{E}(t) \; .
\end{aligned}$$

The approximation holds in the limit that the exponential variation is negligible over the range of the resolution function, *i.e.* $\sigma_t \ll \tau$. A phase shift of the order of $w\sigma^2/\tau$ has also been neglected. The error function turn-on is taken to be superseded by that of the efficiency function. Accordingly, the resolution effects approximately reduce to a re-scaling of the cosine factor – that is to say, to a decrease of the oscillation amplitude, by

$$D_l(w, \sigma_t) \;\; = \;\; e^{-\frac{\sigma_t^2 w^2}{2}} \; . \tag{11.5}$$

In general, the decay distance uncertainty dominates over the momentum uncertainty, and the latter may then be safely neglected.

Partial reconstruction

In partial reconstructed B systems a correcting κ-factor to the proper decay time needs to be considered, as addressed in Section 5.2.3,

$$t \;\; \mapsto \;\; t \cdot \kappa \; .$$

The signal t PDF is modified accordingly,

$$\int \frac{\kappa}{\tau} e^{-\frac{\kappa t}{\tau}} \theta(t) \big(1 + \xi\mathcal{D}\cos(w\kappa t)\big) \, \mathcal{F}(\kappa)\, d\kappa \; .$$

where $\mathcal{F}(\kappa)$ is the the κ-distribution, defined in Section 5.3.2.

This results effectively in the smearing of the proper decay time, and its action may be estimated as part of the resolution function. For fully reconstructed decays $\mathcal{F}(\kappa) = \delta(\kappa - 1)$ and no effect is introduced. For partially reconstructed decays, we use for the current purpose a Gaussian approximation, based on the average $\bar{\kappa}$ and *rms* deviation σ_κ of the actual κ-distribution. For this latter case we have,

$$\begin{aligned} \mathcal{R}(t',t) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_\kappa} e^{-\frac{(\kappa-\bar{\kappa})^2}{2\sigma_\kappa^2}} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t\bar{\kappa}-t'\kappa)^2}{2\kappa^2\sigma_t^2}} (\frac{1}{\bar{\kappa}})\, d\kappa \\ &\simeq \frac{1}{\sqrt{2\pi(\bar{\kappa}^2\sigma_t^2 + \sigma_\kappa^2 t'^2)}} e^{-\frac{\bar{\kappa}^2(t-t')^2}{2(\bar{\kappa}^2\sigma_t^2+\sigma_\kappa^2 t'^2)}} \end{aligned} \tag{11.6}$$

Accordingly, the κ-factor effect is translated effectively as a time-dependent resolution, analogous to a direct momentum uncertainty contribution as addressed above. It results in a damping of the oscillation, approximated to the form of (11.5). Integrating over the time dependence we obtain

$$\begin{aligned} D_\kappa(w,\sigma_\kappa) &= \int_0^{+\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} e^{-\frac{w^2(\sigma_\kappa t)^2}{2}}\, dt \\ &= \sqrt{\pi} Y e^{Y^2} \mathrm{Erfc}(Y) \qquad \text{with} \quad Y \equiv \frac{1}{\sqrt{2}\sigma_\kappa w\tau}\,. \end{aligned} \tag{11.7}$$

11.1.2 Fourier transform approach

The Fourier Transform (FT) of a function $f(t)$ is defined as

$$\tilde{f}(\nu) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\nu t}\, dt\,. \tag{11.8}$$

For practical numerical calculations the discrete Fourier transform is more suitable. Algorithms such as the Fast Fourier Transform are particularly computationally efficient (specifically, by reducing the number of needed arithmetical operations from $\mathcal{O}(N)$ to $\mathcal{O}(NlogN)$). For a set of measurements $\{t_i\}_{i=1}^N$ the discrete transformation is defined simply as

$$g(\nu) = \sum_{i=1}^{N} e^{-i\nu t_i}\,. \tag{11.9}$$

If the sample measurements $\{t_i\}$ correspond to a random extraction over a probability distribution function $f(t)$, then $g(\nu)$ is on average proportional to the continuous transform of $f(t)$,

$$\langle g(\nu)\rangle = \sum_{i=1}^{N} \int_{-\infty}^{+\infty} f(t_i)\, e^{-i\nu t_i}\, dt_i = N\, \tilde{f}(\nu)\,.$$

The Fourier transform of the mixing PDF expression (11.1) is given by

$$\tilde{f}_\xi(\nu) \;=\; \frac{N}{2}\left(f_s[\frac{1}{1+i\nu\tau}+\xi\mathcal{D}\frac{1+i\nu\tau}{w^2\tau^2+(1+i\nu\tau)^2}]+(1-f_s)(1+\xi\mathcal{D}_B)\widetilde{B}(\nu)\right) \quad (11.10)$$

and its real part by

$$\begin{aligned}\mathrm{Re}\,\{\tilde{f}_\xi(\nu)\} \;&=\; \frac{N}{2}\big(f_s[\frac{1}{1+\nu^2\tau^2}+\xi\frac{\mathcal{D}}{2}\left(\frac{1}{1+(\nu-w)^2\tau^2}+\frac{1}{1+(\nu+w)^2\tau^2}\right)]\\ &\quad +(1-f_s)(1+\xi\mathcal{D}_B)\mathrm{Re}\,\{\widetilde{B}(\nu)\}\big)\;. \end{aligned} \quad (11.11)$$

Asymmetry

The difference between the expressions for events tagged as unmixed and mixed is given by

$$\tilde{f}_+(\nu)-\tilde{f}_-(\nu) \;=\; N\left(f_s\mathcal{D}\frac{1+i\nu\tau}{w^2\tau^2+(1+i\nu\tau)^2}+(1-f_s)\mathcal{D}_B\widetilde{B}(\nu)\right)\;.$$

Its real part evaluated at the signal oscillation frequency is

$$\begin{aligned}\Delta(w)\equiv\mathrm{Re}\,\{\tilde{f}_+(w)-\tilde{f}_-(w)\} \;&=\; N\left(f_s\mathcal{D}\frac{1+2w^2\tau^2}{1+4w^2\tau^2}+(1-f_s)\mathcal{D}_B\widetilde{B}(w)\right)\\ &\simeq\; \frac{1}{2}Nf_s\mathcal{D}\;. \end{aligned} \quad (11.12)$$

The approximation is very good for large oscillation frequencies, $w^2\tau^2\gg 1$. For the background, the following considerations need to be made. In general the background flavor asymmetry is quite small, $|\mathcal{D}_B|\ll 1$. Furthermore, long lived background components have a negligible contribution, as long as $w^2\tau_B^2\gg 1$. For less common situations, where sizable short-lived or prompt background components present a large asymmetry, the background cancellation in (11.12) becomes incomplete, and deviations are expected, especially for lower frequencies. Possible uncertainties on the description of such components require that corresponding systematic uncertainties be adequately evaluated.

Detector resolution

The effect of the proper decay time resolution was described by the convolution with the resolution function. Making use of the convolution theorem, $\widetilde{f\otimes g}=\tilde{f}\cdot\tilde{g}$, this is expressed as a multiplicative factor corresponding to the FT of the Gaussian resolution function,

$$D_l(\nu,\sigma_t) \;=\; e^{-\frac{\sigma_t^2\nu^2}{2}}\;.$$

Partial reconstruction

The effects of partial momentum reconstruction are translated into frequency smearing of the FT, for the signal component, via the κ-factor distribution,

$$\begin{aligned}\tilde{f}_\xi(\nu) &= \frac{N}{2}\big(f_s\int_0^\infty [\frac{1}{1+i\nu\tau/\kappa}+\xi\mathcal{D}\frac{1+i\nu\tau/\kappa}{w^2\tau^2+(1+i\nu\tau/\kappa)^2}]\,\mathcal{F}(\kappa)\,d\kappa \\ &+ (1-f_s)(1+\xi\mathcal{D}_B)\widetilde{B}(\nu)\big)\ .\end{aligned}$$

The asymmetry becomes

$$\Delta(w) = Nf_s\mathcal{D}\int_0^\infty \frac{1+w^2\tau^2(1+1/\kappa^2)}{1+2w^2\tau^2(1+1/\kappa^2)+w^4\tau^4(1-1/\kappa^2)^2}\,\mathcal{F}(\kappa)\,d\kappa$$

In case the κ-distribution is well localized around unity, $\mathcal{F}(\kappa)\sim\delta(\kappa-1)$, this expression reduces to (11.12). As its mean shifts to lower values, becoming broader, the term $(1-1/\kappa^2)$ stops being suppressed, and a damping of the signal peak is induced, more prominently so at larger oscillation frequencies.

The smearing effect caused by partial momentum reconstruction is more readily estimated by considering its effective contribution to the resolution of the proper decay time of the B meson. This contribution (11.6) has a time dependence, and the calculation cannot be simply performed using the convolution theorem as above (11.13). Instead one has

$$\begin{aligned}\widetilde{a\otimes g_p}\,(\nu) &= \int_0^{+\infty}\frac{1}{\tau}e^{-\frac{t'}{\tau}}\cos(wt')\left(\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_p t'}e^{-\frac{1}{2}\left(\frac{t-t'}{\sigma_p t'}\right)^2}e^{-i\nu t}\,dt\right)dt' \\ &= \int_0^{+\infty}\frac{1}{\tau}e^{-\frac{t}{\tau}}\cos(wt)e^{-\frac{\nu^2\sigma_p^2t^2}{2}}e^{-i\nu t}\,dt\ .\end{aligned}$$

This determines a reduction of the peak amplitude by a factor approximately equal to $D_p(w,\sigma_p)$ given in (11.7).

Proper time bias

The sculpting of the proper decay time distribution induced by possible trigger and selection criteria is accounted for by the t-efficiency function. The later is to be applied to the reconstructed proper time, thus after the resolution effects have been incorporated. This introduces complexity in the computation to be performed in frequency space. However, these are identical to those appearing in the context of PDF normalization in time space, as addressed in Appendix 13. Alternatively, however, the FT can be evaluated using the inverse of the convolution theorem, which gives

$$\widetilde{f\mathcal{E}} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\tilde{f}(\nu-\lambda)\,\widetilde{\mathcal{E}}(\lambda)\,d\lambda\ .$$

The FT of the indicated typical form of the t-efficiency function is given by

$$\mathcal{E}(t) = \sum_n a_n (t-\zeta_n)^n e^{-\frac{t}{\tau_n}} \theta(t-\zeta_n) \qquad \mapsto \qquad \widetilde{\mathcal{E}}(\nu) = \sum_n a_n \frac{n!\,\tau_n^{n+1}}{(1+i\nu\tau_n)^{n+1}} e^{-(1+i\nu\tau_n)\frac{\zeta_n}{\tau_n}} \ .$$

where the integral computation is reduced to the Gamma function.

The qualitative effect of a proper time bias can be obtained using the simplest form for the t-efficiency curve, namely a direct cut $t > \zeta$ as in (5.17). In the absence of resolution effects,

$$\begin{aligned} &\mathrm{Re}\,\{\int_0^{+\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} \cos(wt)\theta(t-\zeta) e^{-i\nu t}\} \simeq \mathrm{Re}\,\{e^{-\frac{\zeta}{\tau}} \frac{e^{-i(\nu-w)\zeta}}{1+i(\nu-w)\zeta}\} \\ &\simeq \frac{e^{-\frac{\zeta}{\tau}}}{1+(\nu-w)^2\tau^2} \left(\cos\left((\nu-w)\zeta\right) - \sin\left((\nu-w)\zeta\right) \cdot (\nu-w)\tau\right) \ . \end{aligned} \tag{11.13}$$

The amplitude of the signal peak decreases by a factor of $e^{-\frac{\zeta}{\tau}}$, as observed for $\nu = w$. A less anticipated effect, perhaps, is the modification induced in the FT profile. If no bias is present, *i.e.* $\zeta = 0$, the shape is that of the standard Breit-Wigner, $1/(1+(\nu-w)^2\tau^2)$. For non-negligible bias, $\zeta > 0$, however, a sinusoidal behavior is introduced. It results in particular, for realistic parameter values, in a small undershooting of the FT (or Amplitude) profile immediately before and after the signal peak. Note also that the sinusoidal terms also cause the narrowing of the signal peak. Further considerations about these effects are provided in Appendix .5.

Statistical noise

Statistical fluctuations of the sample generate noise in the FT. While the peak size of a signal is given by (11.12), the determination of its statistical significance requires that an evaluation of the noise fluctuations be performed.

The real part of the discrete FT (11.9) is

$$\mathrm{Re}\,\{g(\nu)\} \ = \ \sum_j \cos(\nu t_j) \ .$$

Its variance is evaluated as

$$\begin{aligned}
\sigma^2\big(\mathrm{Re}\,\{g(\nu)\}\big) &= \langle \mathrm{Re}\,\{g(\nu)\}^2\rangle - \langle \mathrm{Re}\,\{g(\nu)\}\rangle^2 \\
&= \sum_{j,k\neq j} \langle \mathrm{Re}\,\{e^{-i\nu t_j}\}\rangle \cdot \langle \mathrm{Re}\,\{e^{-i\nu t_k}\}\rangle + \sum_{j,k=j} \langle \mathrm{Re}\,\{e^{-i\nu t_j}\}\cdot \mathrm{Re}\,\{e^{-i\nu t_k}\}\rangle \\
&\quad - \sum_{j,k} \langle \mathrm{Re}\,\{e^{-i\nu t_j}\}\rangle \cdot \langle \mathrm{Re}\,\{e^{-i\nu t_k}\}\rangle \\
&= (N^2-N)\langle \mathrm{Re}\,\{e^{-i\nu t_j}\}\rangle^2 + N\langle \mathrm{Re}\,\{e^{-i\nu t_j}\}^2\rangle - N^2\langle \mathrm{Re}\,\{e^{-i\nu t_j}\}\rangle^2 \\
&= \frac{N}{2}\left(1 - 2\left(\frac{\mathrm{Re}\,\{\tilde f(\nu)\}}{\tilde f(0)}\right)^2 + \frac{\mathrm{Re}\,\{\tilde f(2\nu)\}}{\tilde f(0)}\right), \qquad (11.14)
\end{aligned}$$

where we have used

$$\begin{aligned}
\langle \mathrm{Re}\,\{e^{-i\nu t}\}\rangle &= \frac{\mathrm{Re}\,\{\int e^{-i\nu t} f(t)\,dt\}}{\int f(t)\,dt} = \frac{\mathrm{Re}\,\{\tilde f(\nu)\}}{\tilde f(0)} \\
\langle \mathrm{Re}\,\{e^{-i\nu t}\}^2\rangle &= \frac{\int \cos^2(\nu t) f(t)\,dt}{\int f(t)\,dt} = \frac{\int (1+\cos(2\nu t))\, f(t)\,dt}{2\,\tilde f(0)} = \frac{1}{2} + \frac{1}{2}\frac{\mathrm{Re}\,\{\tilde f(2\nu)\}}{\tilde f(0)}\,.
\end{aligned}$$

We now apply the results of (11.14) to the PDF of (11.10). For frequencies in the vicinity of the true oscillation frequency w, the real part of the Fourier transform (11.11) and its normalization become

$$\begin{aligned}
\mathrm{Re}\,\{\tilde f_\xi(w)\} &= \frac{N}{2}\big(f_s[\frac{1}{1+w^2\tau^2} + \xi\mathcal{D}\frac{1+2w^2\tau^2}{1+4w^2\tau^2}] + (1-f_s)(1+\xi\mathcal{D}_B)\mathrm{Re}\,\{\widetilde B(w)\}\big) \\
&\simeq \frac{N}{4} f_s \xi\mathcal{D} \qquad (w^2\tau^2, w^2\tau_B^2 \gg 1) \\
\mathrm{Re}\,\{\tilde f_\xi(2w)\} &= \frac{N}{2}\big(f_s[\frac{1}{1+w^2\tau^2} + \xi\frac{\mathcal{D}}{2}(\frac{1}{1+w^2\tau^2} + \frac{1}{1+9w^2\tau^2})] \\
&\quad +(1-f_s)(1+\xi\mathcal{D}_B)\mathrm{Re}\,\{\widetilde B(w)\}\big) \simeq 0 \qquad (w^2\tau^2, w^2\tau_B^2 \gg 1) \\
\tilde f_\xi(0) &= \frac{N}{2}\big(f_s[1+\xi\mathcal{D}\frac{1}{1+w^2\tau^2}] + (1-f_s)(1+\xi\mathcal{D}_B)\big) \\
&\simeq \frac{N}{2} \qquad (w^2\tau^2 \gg 1, |\mathcal{D}_B| \ll 1)\,.
\end{aligned}$$

It can be seen that for large frequencies the ratio of the FT real part to the normalization becomes much smaller than unity. We have then for the uncertainty

$$\begin{aligned}
\sigma^2\big(\mathrm{Re}\,\{\tilde f_\xi(w)\}\big) &= \frac{N_\xi}{2}\left(1 - 2\left(\frac{\mathrm{Re}\,\{\tilde f_\xi(\nu)\}}{\tilde f_\xi(0)}\right)^2 + \frac{\mathrm{Re}\,\{\tilde f_\xi(2\nu)\}}{\tilde f_\xi(0)}\right) \\
&\simeq \frac{N_\xi}{2}\left(1 - \frac{f_s^2\mathcal{D}^2}{2}\right) \qquad (w^2\tau^2 \gg 1)
\end{aligned}$$

where N_ξ denote the number of events with the corresponding tagging decision. It is given by

$$N_\xi \;=\; \tilde{f}_\xi(0) \qquad \text{with} \quad N_+ + N_- = \tilde{f}_+(0) + \tilde{f}_-(0) = N \; .$$

Finally, we arrive at

$$\begin{aligned} \sigma(\Delta) \;&=\; \sqrt{\sigma^2\big(\mathrm{Re}\,\{\tilde{f}_+(w)\}\big) + \sigma^2\big(\mathrm{Re}\,\{\tilde{f}_-(w)\}\big)} \\ &\simeq\; \sqrt{\frac{N}{2}}\sqrt{1-\frac{f_s^2\mathcal{D}^2}{2}} \end{aligned} \tag{11.15}$$

The result in (11.15) is obtained also in [33], through application of the Wiener-Kinchin theorem for the *rms* power spectrum. It is seen that the variance in the FT, from statistical fluctuations, has a behavior independent of the frequency. It is also important to note that all events in the sample, including background, contribute, and that this is to leading order irrespective of resolution and mistagging.

Significance

We have seen that the size of the FT resonance peak is given by

$$\Delta \;=\; \frac{1}{2} N f_s \,\mathsf{D}$$

where D stands for the combined dilution factors

$$\mathsf{D} \;=\; \mathcal{D}\, D_l(w,\sigma_t)\, D_p(w,\sigma_p)$$

from tagging $\mathcal{D}$, vertex resolution D_l, and effective momentum resolution D_p. The noise contribution (11.15) is

$$\sigma(\Delta) = \sqrt{\frac{N}{2}}\sqrt{1-\frac{f_s^2\mathsf{D}^2}{2}} \simeq \sqrt{\frac{N}{2}} \; .$$

The statistical significance of the signal peak is finally

$$\mathcal{S} \;=\; \frac{\Delta}{\sigma(\Delta)} = \sqrt{\frac{N}{2}} f_s \,\mathsf{D} \; . \tag{11.16}$$

We now denote the total number of signal and background events by S and B, respectively. We also notice that we've been considering in the above calculations only flavor tagged candidates. Denoting the tagging efficiency by ϵ, we have $N = \epsilon(S+B)$ We further note that the significance expression (11.16) constitutes an estimate of the uncertainty $\sigma_{\mathcal{A}}$ of the amplitude [33]. It may accordingly be re-written as

$$\frac{1}{\sigma_{\mathcal{A}}} = \sqrt{\frac{\epsilon\mathcal{D}^2\, S}{2}}\sqrt{\frac{S}{S+B}}\,\exp\left(-\frac{\sigma_t^2 w^2}{2}\right)\, D_p(w,\sigma_p) \; . \tag{11.17}$$

11.2 Sensitivity estimation

The effect on the mixing analysis sensitivity of improvements, such as those related to the increases in the tagging power and in the size of the collected data samples, may be estimated in an analytical fashion, as derived in Section 11.1. Sensitivity estimates may thus be obtained through the re-scaling of the resulting significance curve.

The factors which determine the statistical uncertainty on the amplitude, $\sigma_{\mathcal{A}}$, and therefore the sensitivity of the data, include the following:

- the signal yield S and purity S/B; the dependence is $\sigma_{\mathcal{A}} \propto \frac{\sqrt{S+B}}{S}$. These factors depend on the delivered luminosity, the data acquisition and trigger systems efficiency, and the performance of signal extraction.

- the tagging power, $\epsilon\mathcal{D}^2$; the dependence is given by $\sigma_{\mathcal{A}} \propto \frac{1}{\sqrt{\epsilon\mathcal{D}^2}}$. It corresponds to the performance of the applied flavor tagging methods.

- the B proper decay time resolution, σ_{t_B}; the dependence is $\sigma_{\mathcal{A}} \propto e^{(\sigma_{t_B} w)^2/2}$. This corresponds to the accuracy of the measurement of the B proper decay time t_B, given by the decay length and momentum resolutions. It is determined by the experiment's tracking systems, the track and the vertex reconstruction algorithms, and, for the partially reconstructed modes, by how well the B momentum can be inferred. Due to the exponential dependency referred, this factor becomes increasingly more determining when larger oscillation frequencies w are probed.

The statistical uncertainty expected for the fitted amplitude is accordingly expressed by (11.17). In the case of fully reconstructed decays, this becomes

$$\sigma_{\mathcal{A}} \sim \sqrt{\frac{2}{\epsilon\mathcal{D}^2}} \cdot \frac{\sqrt{S+B}}{S} \cdot \exp\left(\frac{\sigma_t^2 w^2}{2}\right) \qquad \text{(exclusive modes)} \,,$$

where the proper decay time resolution is determined by the vertex resolution. For partially reconstructed decays, an additional factor D_p, expressed in (11.7), needs to be considered, due to the uncertainty in the B momentum determination caused by the incomplete reconstruction,

$$\sigma_{\mathcal{A}} \sim \sqrt{\frac{2}{\epsilon\mathcal{D}^2}} \cdot \frac{\sqrt{S+B}}{S} \cdot \exp\left(\frac{\kappa^2\sigma_t^2 w^2}{2}\right) \cdot D_p^{-1}(w, \sigma_p) \qquad \text{(inclusive modes)} \,,$$

where σ_p is estimated as the ratio of the root-mean-square deviation to the average of the κ -factor distribution $\mathcal{F}(\kappa)$ (Section 5.3.2), which are about 0.12 and 0.80, respectively.

Representative average values for the hadronic and semileptonic B_s samples are illustrated in Table 11.1. The evaluation of the significance based on the analytical expressions

	hadronic	semileptonic
signal yield, S	1100	15000
signal-to-noise ratio, S/B	3.4	2.3
vertex resolution, σ_t [fs]	96	157
momentum resolution, σ_p	–	0.15

Table 11.1: Representative values of B_s sample parameters.

above, however, is performed individually for each decay mode. It takes into account the distributions of proper decay time resolution σ_t. These distributions are obtained from mass sideband subtracted data, after calibration (Section 5.7), and are illustrated in Figure 5.1. The κ-factor distributions, illustrated in Figure 5.4, are also employed. The different contributions, as given by those equations, are added in quadrature, as it is done when combining the contributions from the various decay channels.

The amplitude uncertainties $\sigma_{\mathcal{A}}(w)$ measured in the hadronic and semileptonic data samples, and from the combined scans, shown in Figures 8.5 – 8.7 initially obtained with the OST applied to the $355\,\mathrm{pb}^{-1}$ dataset in Chapter 8, are reproduced in Figure 11.1. The corresponding analytical significance curve projections are also represented, and are seen to provide an appropriate description. The analytical curves have an intrinsic smooth behavior, which replaces the statistical fluctuations that characterize the measured values, especially at higher probed frequencies. Also, they further provide an extension beyond the probed spectrum range. The curves predicted with the inclusion of the SST are also overlaid in Figure 11.1. The tagging power values used are those reproduced in Table 11.2. The estimated sensitivity values are given by the intersection of those curves with the horizontal unit line also displayed.

OST $\epsilon\mathcal{D}^2$	0.0155
SST $\epsilon\mathcal{D}^2$	0.04
combined $\epsilon\mathcal{D}^2$	0.0455

Table 11.2: Input tagging effectiveness.

The procedure employed for sensitivity estimation can be summarized as follows:

1. check the analytical description against measured $\sigma_{\mathcal{A}}$ values,

2. re-generate $\sigma_{\mathcal{A}}$ curves according to improvements considered,

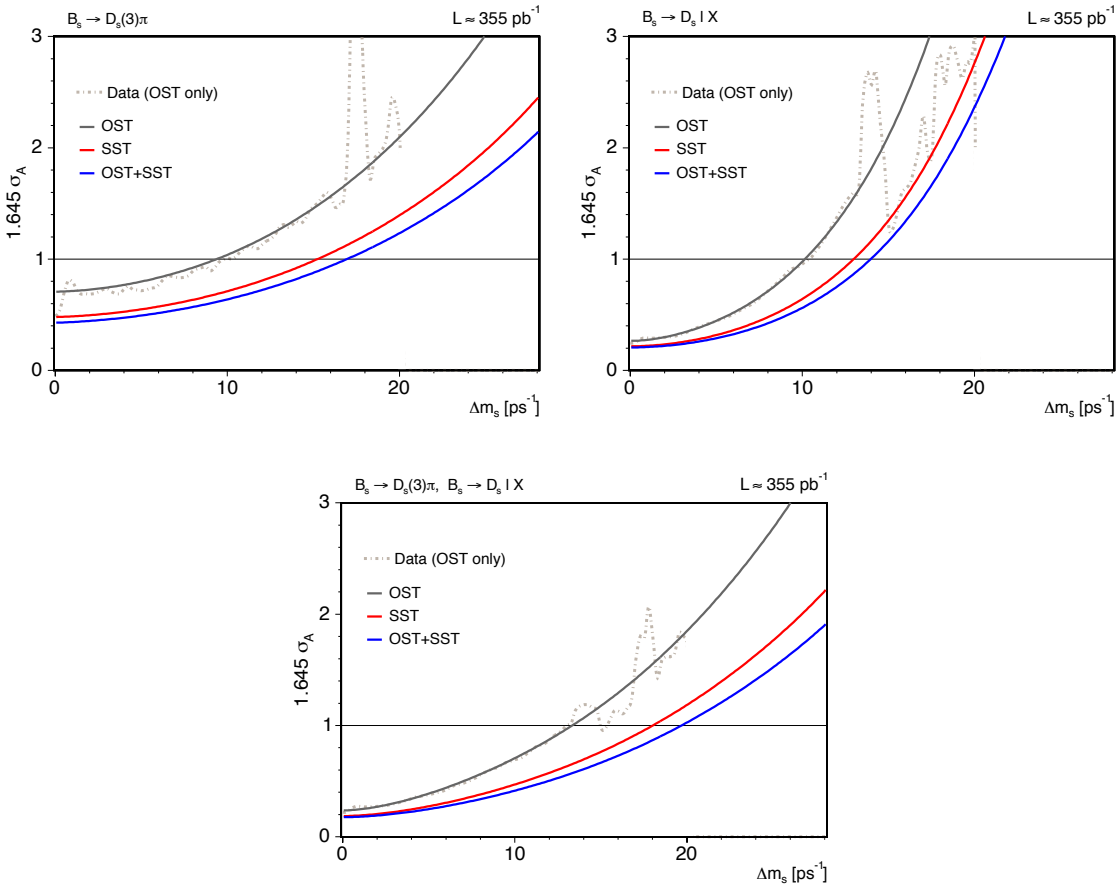

Figure 11.1: Significance curves for the hadronic, semileptonic and combined analyses, with estimated improvements lead by the inclusion of the SST; the histogram represents the observed significance obtained in the analysis of 355 pb^{-1} with only OST taggers which is taken as the starting reference.

3. identify corresponding sensitivity value as intersection point with line of the desired confidence level.

The projected sensitivities are displayed in the graphs of Figure 11.2, as a function of a scale to the yields obtained in the analysis of first 355 pb^{-1} of data. Sensitivities are evaluated for 95% exclusion, as well as 3σ and 5σ observation. It demonstrated that the exclusion of frequency values of the order of 30 ps^{-1} or the observation of a signal as high as 20 ps^{-1} was indeed within reach under reasonably conservative assumptions. It also shows that the fully reconstructed decay samples tend to provide the leading contribution, as anticipated, for such large frequencies, while the semileptonic samples contribute an additional meaningful contribution to the combined results.

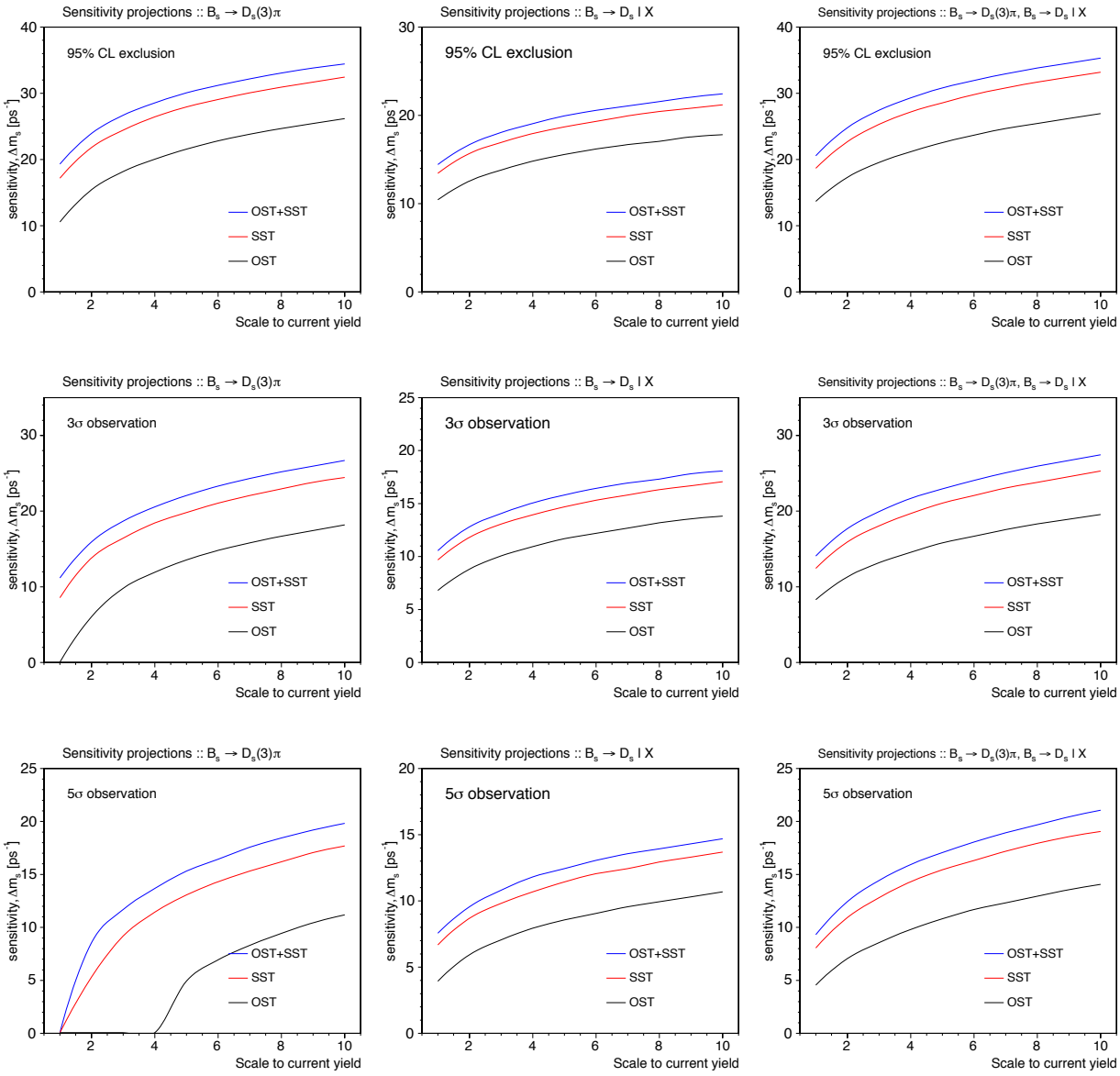

Figure 11.2: Sensitivity extrapolations for the hadronic, semileptonic, and combined analyses; estimates (y-axis) are shown as a function of increasing signal yields (x-axis), represented relatively to those extracted using $355\,\mathrm{pb}^{-1}$ (*current*), and are provided for the confidence levels of 95% exclusion, 3σ and 5σ observation; the individual contributions from OST and SST are also displayed.

11.3 Probability of random fluctuation

The amplitude significance, $\mathcal{A}/\sigma_{\mathcal{A}}$, has been tacitly employed also as a measure of the oscillation signal significance. In Chapter 10 we have used this estimator, evaluated at a single frequency point, as a comparative indicator of significance improvements. Here however we are to ask a more general question which is of crucial interest specifically in declaring the measurement an *observation*. Specifically, we address now the issue of quantifying the probability for a random fluctuation to produce a signal of the oscillations as large or larger than the one observed in the data.

In statistical significance testing, such a measure is referred to as the *p-value*, and is defined as the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The test statistic adopted is the minimum $\mathcal{R}_{\min}$ of the log-likelihood ratio $\mathcal{R}$ defined as

$$\mathcal{R} = -\ln \frac{\mathcal{L}^{\mathcal{A}=1}(\Delta m_s)}{\mathcal{L}^{\mathcal{A}=0}} \,. \tag{11.18}$$

This relates the likelihood calculated with $\mathcal{A} = 1$, which corresponds to the hypothesis that Δm_s is the true mixing frequency, and with $\mathcal{A} = 0$, which is independent of Δm_s and corresponds to the null hypothesis of absence of oscillations.

The measured likelihood is shown in Figure 11.3 for the final combined B_s analysis presented in Section 10.3. As can be seen, the statistical fluctuations of the normalized likelihood profile at high probe frequencies are increasingly small about zero and do not approach the depth of the signal minimum. Note in contrast that statistical fluctuations do not have an identical effect on the $\mathcal{A}/\sigma_{\mathcal{A}}$ estimator, as they cause it to be distributed as a unit Gaussian about zero, including some large values. The $\mathcal{R}_{\min}$ estimator, going beyond single-point estimates or limited range scans, represent a comprehensive p-value indicator for an extended frequency range.

The probability density function, $f(\mathcal{R}_{\min})$, of the test statistic $\mathcal{R}_{\min}$ is estimated directly from data, by randomizing the tagging decision many times independently. The measured $\mathcal{R}_{\min}$ distribution, obtained from about 350 million randomized tag iterations, is shown on the left plot of Figure 11.4, which upon unit normalization is taken as the $f(\mathcal{R}_{\min})$ estimate. The frequency range of the amplitude scans performed on the data is arbitrarily chosen to be from 0 to 35 ps^{-1}. The estimation does not incurr any significant bias by selecting a finite probe frequency window, given that the ratio converges rapidly to zero at higher frequency values. An extension of the search range up to 50 ps^{-1} was performed, as a cross-check, with no resulting effect on the $f(\mathcal{R}_{\min})$ distribution. The cumulative distribution of the test statistic, shown in the right plot of Figure 11.4, yields the calculated p-value as a function of

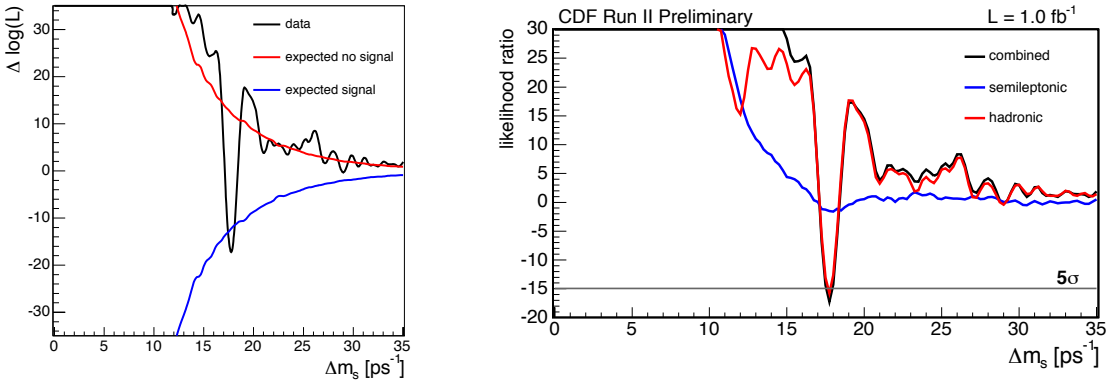

Figure 11.3: Normalized likelihood profile, denoted $\mathcal{R}$ in the text, for the B_s oscillation observation, and comparison to the mixing and no-mixing expectations, shown for the full probed frequency spectrum.

$\mathcal{R}_{\rm min}$:

$$p(\mathcal{R}_{\rm min}^{observed}) = \int_{-\infty}^{\mathcal{R}_{\rm min}^{observed}} f(\mathcal{R}_{\rm min})\, d\mathcal{R}_{\rm min}\,. \tag{11.19}$$

The observed likelihood ratio minimum, in the final analysis of Section 10.3, is $\mathcal{R}_{\rm min}^{observed} = -17.26$. The corresponding p-value is inferred from (11.19) to be 8×10^{-8}. This is also indicated on Figure 11.4. From the 3.5×10^8 randomized trials, 28 scans were found with a $\mathcal{R}_{\rm min}$ value smaller than -17.26. This p-value corresponds to a Gaussian significance of 5.4 standard deviations. As also shown on the left plot of Figure 11.4, this p-value stands well beyond the five standard-deviations threshold of 5.7×10^{-7}.

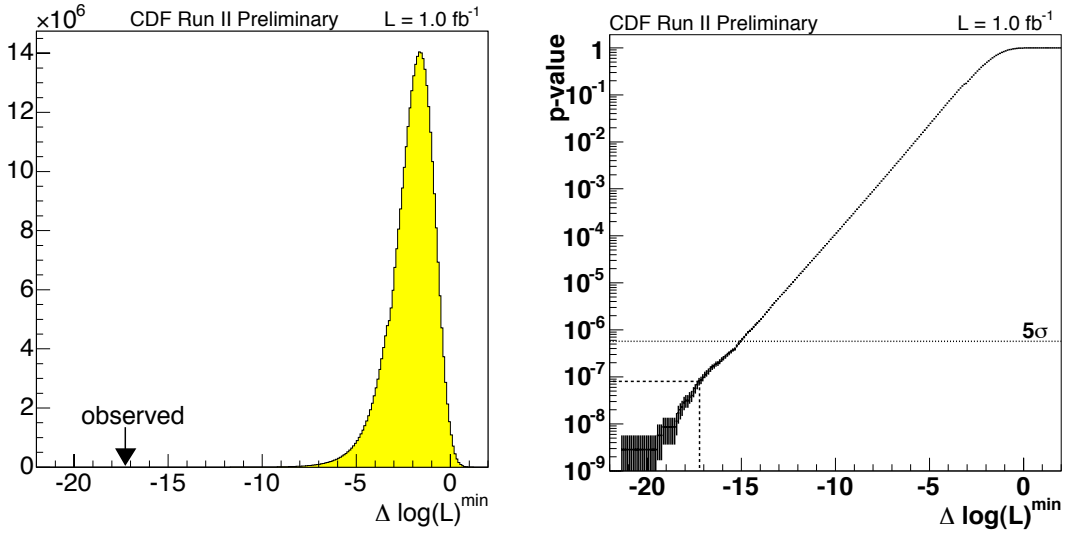

Figure 11.4: Distribution of the likelihood-ratio minima, $\mathcal{R}_{\min}$, for fluctuations of randomized data (left), and resulting p-value dependence; the observed signal result, $\mathcal{R}_{\min} = -17.26$, and corresponding p-value are also indicated.

11.4 Résumé

An analytical digression through the underlying formalism of the mixing analysis has been presented. Important insights are derived from the associated likelihood and Fourier treatments, including a characterization of the relevant experimental effects. The expression for the expected single-point amplitude significance for a mixing signal is derived, and further used to perform sensitivity estimates based on the properties of the mixing data samples.

The mechanism employed to assess the probability of random fluctuations in the data to mimic the oscillation signal is detailed. The probability that randomly-tagged data would produce a signal as large as the one observed in the final analysis, reported in Section 10.3, is found to be beyond 5σ.

This serves as confirmation of the analysis as the world's first definitive *observation* of B_s oscillations.